

# Better data with fewer participants and trials: Improving experiment efficiency with adaptive design optimization

Daniel R. Cavagnaro (cavagnaro.2@osu.edu)

Department of Psychology, 225 Psychology Building, 1835 Neil Avenue  
Columbus, OH 43210 USA

Yun Tang (tang.162@osu.edu)

Jay I. Myung (myung.1@osu.edu)

Mark A. Pitt (pitt.2@osu.edu)

## Abstract

The design of an experiment can greatly affect its potential to produce statistically conclusive results. In this paper, we offer a method for increasing the statistical informativeness of an experiment through the use of adaptive design optimization. The problem to be solved in adaptive design optimization is identifying an experimental design under which one can infer the underlying model and its parameter values in the fewest possible steps. While this problem is often impossible to solve analytically, it can be solved numerically with the help of a Bayesian computational trick that recasts it as a probability density simulation in which the optimal design is the mode of the density. The resulting optimization algorithm is flexible enough to apply in a variety of experimental settings. We demonstrate the effectiveness of this approach with a computer simulation of a memory retention experiment.

**Keywords:** Design optimization; Model discrimination; Computer simulation; Retention

## Introduction

Experimentation is fundamental to the advancement of science, whether one is interested in studying the neuronal basis of a sensory process or assessing the efficacy of a new drug in clinical trials. Often in cognitive science experiments, the data collected in an experiment is used to fit some set of statistical models of the process under investigation. Before data collection can even begin, however, many choices about the design of the experiment must be made. In particular, design parameters such as the sample size, the number of treatments (i.e., conditions or levels of the independent variable) to study, and the proportion of observations to be allocated to each treatment group must be chosen. These choices impact not only the statistical value of the results, but also the cost of the experiment. For example, basic statistics tells us that increasing the sample size would increase the statistical power of the experiment, but it would also increase the cost (e.g., number of participants, amount of testing). An optimal experimental design is one that maximizes the informativeness of the experiment, while being cost effective for the experimenter.

To see how an optimal design might look in practice, consider the following example of a typical experiment designed to discriminate models of retention (i.e., forgetting). The experiment consists of a ‘study phase,’ in which subjects are given a list of words to memorize, followed by a time interval, followed by a ‘test phase,’ in which subjects are quizzed

Model	Equation
Power (POW)	$p = a(t + 1)^{-b}$
Exponential (EXP)	$p = ae^{-bt}$

Table 1: Two statistical models of memory retention. In each equation, the symbol  $p$  ( $0 < p < 1$ ) denotes the predicted probability of correct recall as a function of time interval  $t$  with model parameters  $a$  and  $b$ .

to see how many words they can recall from the list. The number of words recalled correctly after each time interval can then be plotted on a graph to reveal how retention changes over time. A model of retention is essentially the functional form of a curve that can fit this relationship between retention and time.

Two models that have been of interest are the power model (POW) and the exponential model (EXP), both of which are defined in Table 1 (Rubin & Wenzel, 1996). Their predictions for a very narrow range of parameters are depicted in the first column of Figure 1 (predictions are presented as probabilities, with darker regions corresponding to higher values). Both models predict that memory decays monotonically as the time interval between the study phase and the test phase increases, but forgetting is predicted to occur much more quickly immediately after the study phase in the power model than in the exponential model. If a researcher conducts an experiment to compare the predictions of these models, and has the resources to probe memory at up to five time intervals between the range of 0 and 20 seconds after study, then the goal of design optimization is to identify those time intervals that yield the most differentiating predictions of the two models. Visual inspection of Figure 1 suggests that the values should fall between 2 and 7 seconds, where the prediction of the power model is most different from of the exponential model. For example, as can be seen in first column of probability scatter plots in Figure 1, the power model predicts between 10 and 20 correct responses at a lag time of 5 seconds, with around 15 being the most likely outcome, whereas the exponential model predicts between 25 and 45 correct responses at 5 seconds with around 35 being the most likely outcome.

This simple example of finding an optimal design by visual

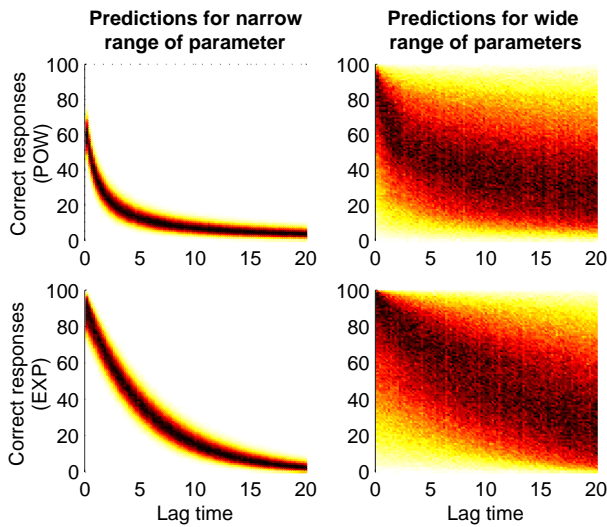


Figure 1: Scatter plot of the probabilities of each number of correct responses, out of 100 Bernoulli trials, predicted by power model (top row) and exponential model (bottom row) for a narrow range of parameters (left column) and a wide range of parameters (right column). Darker colors represent higher probabilities.

inspection is meant to illustrate the more complicated problems that arises in practice. In actual retention experiments, it would be highly unusual to have enough prior information to restrict the model parameters to such a small range. When the parameters are allowed to vary across a more realistic range, the picture looks more like the second column of Figure 1. In that case, an optimal design for discriminating the two models can not be found by visual inspection alone.

Given how closely the power and exponential models can mimic one another when their parameters are allowed to vary realistically, finding an optimal design for distinguishing them is no easy task. A principled approach to this challenging problem can be found in Bayesian decision theory (Chaloner & Verdinelli, 1995). In the Bayesian framework, each possible experimental design is treated as a gamble whose payoff is determined by the outcome of an experiment carried out with that design. The idea is to estimate the informativeness of hypothetical experiments carried out with a given design, so that an "expected utility" (i.e., informativeness) of that design can be computed. This is done by considering every possible observation that could be obtained from an experiment with that design and then evaluating the relative likelihoods and utilities of those observations. The design with the highest expected utility is then chosen as the optimal design.

The Bayesian approach to experimental design optimization extends easily to the case of adaptively designed experiments, in which testing proceeds over the course of several stages (e.g., trials), and the design of the experiment can be

updated at each stage. At each stage, the information gained from all prior stages is used to improve the design at the current stage. Thus, the problem to be solved in *adaptive design optimization* (ADO) is to identify the most informative design at each stage of the experiment, taking into account the results of all previous stages, so that one can infer the underlying model and its parameter values in as few steps as possible.

Finding optimal adaptive designs in a general setting requires simultaneous high-dimensional optimization and integration. This can be an extremely difficult task given that the integration can be intractable for the complex, nonlinear models.

Despite these computational challenges, the flexibility and efficiency of adaptive designs have made them popular in various fields. For example, in astrophysics, ADO has been used in the design of experiments to distinguish between dark energy models (Heavens et al., 2007). ADO has also been used in designing phase I and phase II clinical trials to ascertain the dose-response relationship of experimental drugs (Haines et al., 2003; Ding et al., 2008), as well as in estimating psychometric functions (Kujala & Lukka, 2006; Lesmes et al., 2006). However, because identification of optimal adaptive designs in general settings was not possible until recently, most of this prior work focused on problems that were sufficiently simple (e.g. linear models with normal error) so that analytical solutions could be found.

We draw on a recent breakthrough in stochastic Bayesian optimization (Müller et al., 2004) in offering a method for finding optimal adaptive designs in a general setting, without relying on approximations nor simplifying assumptions. The basic idea is to recast the problem as a probability density simulation in which the optimal design corresponds to the mode of the distribution. This allows one to find the optimal design without having to evaluate the integration and optimization directly. The density is simulated by Markov Chain Monte-Carlo (Gilks et al., 1996), and the mode is sought by gradually "sharpening" the distribution with a simulated annealing procedure (Kirkpatrick et al., 1983).

In the remainder of the paper, we first give a formal account of the Bayesian framework for ADO, including the computational algorithm for finding optimal designs via probability density simulation, and then we demonstrate the implementation of ADO in a computer-simulated retention experiment. Readers interested in only an example of its application should jump to the section "ADO for discriminating retention models." However, we wish to emphasize that this example is just one of many potential applications of ADO. The computational algorithm is very general and works just as well with other types of design variables. Readers interested in additional applications, such as optimization of categorization experiments, should see Myung & Pitt (in press).

## Bayesian Adaptive Design Optimization

Adaptive design optimization within a Bayesian framework has been considered at length in the statistics community (Atkinson & Donev, 1992) as well as in other science and engineering disciplines (e.g. El-Gamal & Palfrey, 1996; Bardley et al., 1996; Allen et al., 2003). The issue is essentially a Bayesian decision problem where, at each stage  $s = (1, 2, \dots)$  of experimentation, the most informative design at stage  $s$  (i.e., that design with the highest expected utility) is chosen based on the outcomes of the previous experiments  $\{y_1, \dots, y_{s-1}\}$ . The criterion for the informativeness of a design often depends on the goals of the experimenter. The experiment which yields the most precise parameter estimates may not be the most effective at discriminating among competing models, for example (see Nelson, 2005, for a comparison of several utility functions that have been used for optimal design of experiments in cognitive science).

Whatever the goals of the experiment may be, solving for the optimal design is a highly nontrivial problem. Formally, Bayesian ADO entails finding an optimal design  $d_s^*$ , at each stage  $s$  of the experiment, which maximizes a utility function  $U_s(d)$  defined as

$$U(d) = \sum_{m=1}^K p(m) \int \int u(d, \theta_m, y) p(y|\theta_m, d) p(\theta_m) dy d\theta_m, \quad (1)$$

where  $m = \{1, 2, \dots, K\}$  is one of a set of  $K$  models being considered,  $d$  is a design,  $\theta_m$  is a parameterization of model  $m$ , and  $y$  is the outcome of an experiment with design  $d$  under model  $m$  with parameter  $\theta_m$ . We refer to the function  $u(d, \theta_m, y)$  in Equation (1) as the *local utility* of the design  $d$ . It measures the utility of a *hypothetical* experiment carried out with design  $d$ , when the true model is  $m$  with parameter  $\theta_m$ , and the sample  $y$  is obtained. Thus,  $U_s(d)$  represents the expected value of the local utility function, where the expectation is taken over all models under consideration, the full parameter space of each model, and all possible outcomes, with respect to the model prior probability  $p_s(m)$ , the prior parameter probability  $p_s(\theta_m)$ , and the sampling distribution function  $p(y|\theta_m)$ , respectively.

The model and parameter priors are updated at each stage  $s = \{1, 2, \dots\}$  of experimentation. Specifically, upon the specific outcome  $z_s$  observed at stage  $s$  of an *actual* experiment carried out with design  $d_s$ , the model and parameter priors to be used to find an optimal design at the next stage are updated via Bayes rule and Bayes factor calculation

$$p_{s+1}(\theta_m) = \frac{p(z_s|\theta_m, d_s) p_s(\theta_m)}{\int p(z_s|\theta_m, d_s) p_s(\theta_m) d\theta_m} \quad (2)$$

$$p_{s+1}(m) = \frac{p_0(m)}{\sum_{k=1}^K p_0(k) BF_{(k,m)}(z_s) p_s(\theta)} \quad (3)$$

where  $BF_{(k,m)}(z_s)$  denotes the Bayes factor defined as the ratio of the marginal likelihood of model  $k$  to that of model  $m$  given the realized outcome  $z_s$ , where the marginal likelihoods are computed using the updated priors from the preceding

stage. The above updating scheme is applied successively on each stage of experimentation, after an initialization with equal model priors  $p_{(s=0)}(m) = 1/K$  and a non-informative parameter prior  $p_{(s=0)}(\theta_m)$ .

## Computational Methods: Finding an optimal design via probability simulation

Given the multiple computational challenges involved in finding the design  $d^*$ , standard optimization methods such as Newton-Raphson and Revelverg-Marquardt are insufficient. However, a promising new approach to this problem has been proposed in statistics (Müller et al., 2004). It is a simulation-based approach that includes an ingenious computational trick that allows one to find the optimal design without having to evaluate the integration and optimization directly in Equation (1). The basic idea is to recast the problem as a simulation from a sequence of augmented probability models.

To illustrate how it works, let us consider the design optimization problem to be solved at any given stage  $s$  of experimentation, and, for simplicity, we will suppress the subscript  $s$  in the remainder of this section. According to the ingenious trick of Müller et al. (2004), we treat the design  $d$  as a random variable and define an auxiliary distribution  $h(d, \cdot)$  that admits  $U(d)$  as its marginal density. Specifically, assuming that  $u(d, \theta_m, y)$  is non-negative and bounded, we define

$$h(d, y_1, \theta_1, \dots, y_K, \theta_K) = \alpha \left[ \sum_{m=1}^K p(m) u(d, \theta_m, y_m) \right] p(y_1, \theta_1, \dots, y_K, \theta_K | d) \quad (4)$$

where  $\alpha(>0)$  is the normalizing constant of the marginal distribution, and

$$p(y_1, \theta_1, \dots, y_K, \theta_K | d) = \prod_{m=1}^K p(y_m | \theta_m, d) p(\theta_m)$$

is the joint prior distribution of observations  $y$  and parameters  $\theta$ . Note that the subscript  $m$  in the above equation refers to model  $m$ , not the stage of experimentation. For instance,  $y_m$  denotes an experimental outcome generated from model  $m$  with design  $d$  and parameter  $\theta_m$ . It can be shown that marginalizing  $h(d, \cdot)$  over  $\{y_1, \theta_1, \dots, y_K, \theta_K\}$  in Equation (4) yields  $h(d) = \alpha U(d)$ . This means that the probability model  $h(d)$  generates designs with probability proportional  $U(d)$ . Consequently, the design with the highest utility according to  $U(d)$  can be found by taking the mode of a sufficiently large sample of designs drawn from the marginal distribution  $h(d)$ .

However, finding the global optimum could potentially require a very large number of samples from  $h(d)$ , especially if there are many local optima or if the design space is very irregular or high-dimensional. To overcome this problem, we augment the auxiliary distribution with independent samples of  $y$ 's and  $\theta$ 's given design  $d$  as follows:

$$h_J(d, \cdot) = \alpha_J \prod_{j=1}^J h(d, y_{1,j}, \theta_{1,j}, \dots, y_{K,j}, \theta_{K,j}) \quad (5)$$

for a positive integer  $J$  and  $\alpha_J(> 0)$ . The marginal distribution of  $h_J(\cdot)$  obtained after integrating out model parameters and outcome variables will then be equal to  $\alpha_J U(d)^J$ . Hence, as  $J$  increases, the distribution  $h_J$  will become more highly peaked around its (global) mode corresponding to the optimal design  $d^*$ , thereby making it easier to identify the mode.

Following Amzal et al. (2006), we implement a sequential Monte Carlo particle filter algorithm that begins by simulating  $h_J(d, \cdot)$  in Equation (5) for  $J = 1$ , and then increases  $J$  incrementally on subsequent iterations on an appropriate simulated annealing schedule (Kirkpatrick et al., 1983; Doucet et al., 2001).

### ADO for discriminating retention models

We conducted a computer simulation to illustrate the ADO procedure, the purpose of which was to design an experiment to discriminate between the retention models in Table 1. The point of the simulated experiment is to demonstrate how quickly an optimal adaptive experiment can correctly identify the true model, which in this case was EXP with parameters  $a = 0.7103$  and  $b = 0.0833$ .

The first step in conducting the simulation was to choose a utility function (i.e., measure of model discriminability) that was appropriate for the goal of the experiment. Since the goal was to discriminate between models POW and EXP, we chose to use a utility function based on the entropy of the posterior model probabilities. Specifically, we set the utility of a design  $d$  to be the mutual information (Cover & Thomas, 1991) between  $M$  and  $Y_d$ , where  $M$  is a random variable representing uncertainty about the true model, and  $Y_d$  is a random variable denoting the result of an experiment carried out with design  $d$ .

The next step was to solve for the optimal design (i.e., the time lags) to use in the initial stage of the experiment, based on the initial (i.e., prior) model and parameter probabilities. We considered designs with one test phase (i.e., recall of the study items) after a single time lag, and a fixed binomial sample size of  $n = 30$  at each stage of experimentation. Using equal prior probabilities on POW and EXP, and independent beta priors on  $a$  and  $b$  in each model<sup>1</sup>, we used the just-described algorithm to find the time lag,  $d_1$ , that maximized  $U_1(d)$ . We then generated data from the true model at  $t = d_1$ , and updated the prior probabilities accordingly. These updated distributions were then used as the priors for finding the optimal design at the next stage, so that the process could be repeated. We continued this process for eight stages of the experiment. A typical profile of the posterior model probability  $p_s(EXP)$  as a function of stage  $s$  is shown by the solid black line in Figure 2. The optimal adaptive design identifies

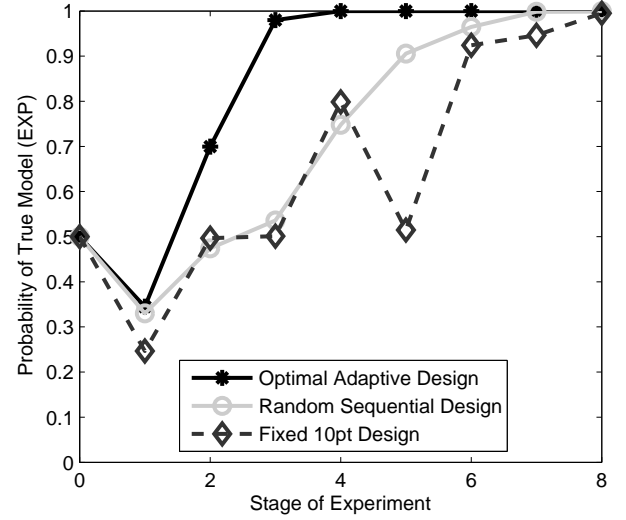


Figure 2: Posterior model probabilities in the simulated experiments.

the correct model with over 0.95 probability after just three stages or 90 Bernoulli observations.

To help illustrate the logic of the ADO procedure in the simulated experiment, the predictions of the two models at each of the first five stages (trials) are shown in Figure 3. Recall that our goal in ADO is to choose, at each stage, the lag time at which the predictions of the models differ the most, and then update model probabilities based on the relative likelihoods of the observed outcome. For example, in stage 2 of the simulated experiment, depicted by the second column of graphs in Figure 3, 96.4 seconds was found to be the optimal design. It can be seen from the stage-2 scatter plots that EXP predicts fewer correct responses than POW at lag times around 96.4 seconds. When zero correct responses were observed (an outcome that is much more likely under EXP than under POW, as indicated by the colors at the tips of the arrows in Figure 3) the odds of EXP are increased from 0.35 to 0.7 in stage 3 and the odds of POW are decreased from 0.65 to 0.3. As this process continues, the predictions of the models narrow around the observed outcomes and the posterior probability of the true model (EXP) approaches 1.

For comparison of ADO against a different sequential design scheme, we also conducted several simulated experiments with randomly generated designs. These randomly designed experiments proceeded in the manner described above, except that the lag time at each stage was chosen randomly between zero and 100 seconds. The solid gray line in Figure 2 shows a typical posterior model probability curve obtained in these experiments. The line shows that the true model could be identified with 0.90 probability after 6 stages or 180 Bernoulli trials; that's nearly twice as many as were required when ADO was used.<sup>2</sup>

<sup>1</sup>We used priors  $a \sim \text{Beta}(2, 1)$  and  $b \sim \text{Beta}(1, 4)$  for the power model and  $a \sim \text{Beta}(2, 1)$  and  $b \sim \text{Beta}(1, 80)$  for the exponential model over each model's parameter space,  $0 < a, b < 1$ . These priors reflect conventional wisdom about these retention models based on many years of investigation. The choice of priors does indeed change the optimal solution, but the importance of this example is the process of finding a solution, not the actual solution itself.

<sup>2</sup>We repeated the simulations several times in order to verify that

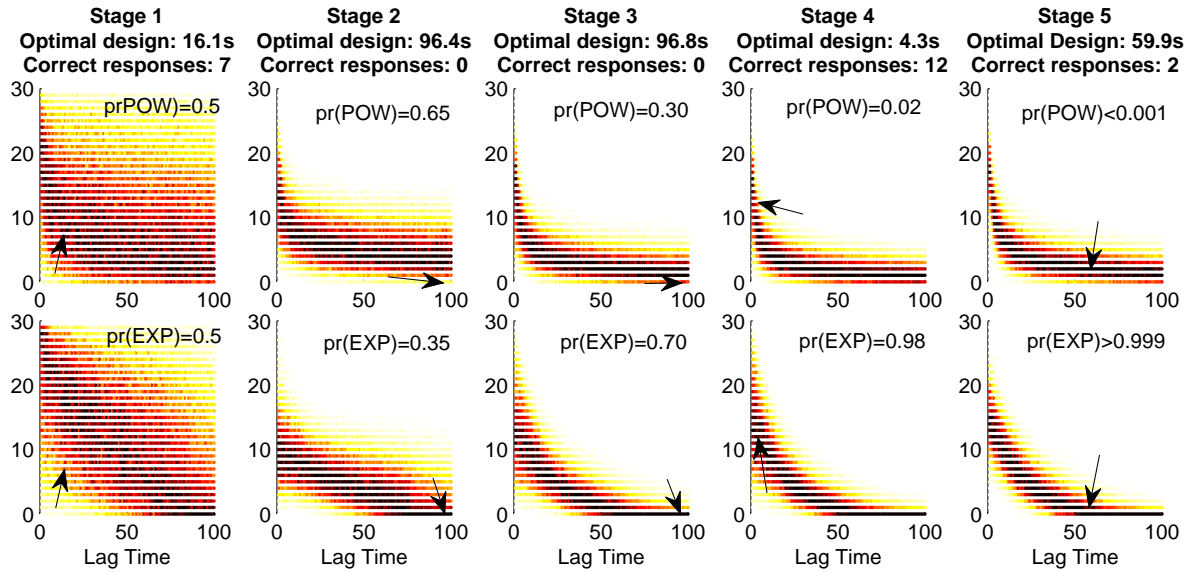


Figure 3: Predictions of the power and exponential models in the first five stages (i.e., trials) of the simulated experiment. The predictions are based on the prior parameter estimates at each stage. For example, the stage 2 column of scatter plots depicts outcome probabilities at each lag time, and for each model, according to parameter estimates that were updated based on the results observed in stage 1. The text above and inside the graphs provide information about the prior probabilities of each model, the optimal designs for discriminating the models, and the observed outcomes (correct responses) at each stage of the simulated experiment. Arrows denote the number of correct responses at the optimal lag time

To compare ADO against a more principled approach than randomly selecting lag times at each stage, we conducted additional simulations using a typical design from the retention literature. While there is no established standard for the set of lag times to test in actual retention experiments, a few conventions have emerged. For one, previous experiments utilize what we call ‘fixed designs,’ in which the set of lag times at which to assess retention are specified before experimentation begins, and held fixed for the duration of the experiment. Thus, there is no Bayesian updating between stages as there would be in a sequential design. The lag times are typically concentrated near zero and spaced roughly geometrically. For example, Rubin et al. (1999) used a design consisting of 10 lag times: (0s, 1s, 2s, 4s, 7s, 12s, 21s, 35s, 59s, 99s). To get a meaningful comparison between this fixed design and the sequential designs that we just tested, we generated data at each stage from the same model as in the previous simulations, but with just 3 Bernoulli trials at each of the 10 lag times in the Rubin et al. design. That way, each stage of the experiment included the same number of total observations as in the experiments with just one lag time at each stage. Posterior model probabilities and parameter estimates were computed after each stage but not carried over as priors for subsequent stages. The obtained posterior model probabilities from a typical simulation are shown by the dashed line in Figure 2.

the results were not due to random chance. Occasionally the random approach stumbled upon a good design and performed very well, and sometimes even an optimal design was hindered by unlikely experimental outcomes, but the reported results reflect typical performances of the two designs.

The results of these simulations clearly demonstrate the advantage of the optimal adaptive design. In the experiment designed with ADO, the true model could be identified with over 0.95 probability after just three stages or 90 Bernoulli observations. In contrast, the experiments that did not use ADO required twice as many observations (6 stages or 180 Bernoulli trials) to produce a similar level of evidence in favor of the true model. This marked improvement in efficiency is even more striking in light of the fact that the ADO experiment was optimized over just one of many possible design variables. Even better designs could potentially be obtained by optimizing over additional design variables, such as the number of lag times to test at each stage, and the number of Bernoulli trials to allocate to each lag time at each stage.

It is worth noting that the preceding simulations are intended only as a proof-of-concept. Obviously, in this simple case, an optimal design could be found via comprehensive grid searches. However, the approach that we have demonstrated here generalizes easily to much more complex problems in which brute-force searchers would be impractical or impossible.

## Other Applications of ADO

The design optimization scheme that we have presented here can be applied in a variety of settings besides the simple retention experiment described above. Simply put, any experiment in which observations can be made sequentially, and for which the goal is to distinguish among closed form mathematical models, could potentially benefit from the applica-

tion of ADO. For a different example, consider the following application to an experiment in the field of cognitive development.

### ADO for Developmental Psychology

It is widely believed that the abstract representation of numbers by children transitions from a logarithmic scale to a linear scale as they grow (Opfer & Siegler, 2007). Verifying this hypothesis leads to the problem of discriminating between linear and logarithmic models in experiments. Typical experiments designed to elicit such numerical representations use a “number-to-position” task, in which children are shown a number between zero and 1000 and asked to estimate its position on a line. The choice of numbers to be shown at each stage of the experiment constitutes a design variable that can be controlled by the experimenter, hence choosing appropriate numbers to show (i.e. designs) in order to quickly and reliably discriminate linear and log models could be seen as a design optimization problem. Simulation results have shown that using ADO to find the optimal set of numbers to show in each trial greatly reduces the number of trials required to reveal the data-generating model (Tang et al., 2009).

### Conclusion

Computational models are precise in their predictions about human behavior. Our experimental methods for probing cognition to assess the accuracy of their predictions are much less so. Adaptive design optimization is a tool that can not only improve the informativeness of an experiment in discriminating models, but also increase the efficiency of data collection. We are currently applying the method in experimentation and expanding its application to more complex experimental designs and models.

### Acknowledgements

This research is supported by National Institute of Health Grant R01-MH57472 to JIM and MAP. We wish to thank Hendrik Kueck and Nando de Freitas for valuable feedback and technical help provided for the project, and Michael Rosner for the implementation of the design optimization algorithm in C++.

### References

- Allen, T., Yu, L., & Schmitz, J. (2003). An experimental design criterion for minimizing meta-model prediction errors applied to die casting process design. *Applied Statistics*, 52, 103-117.
- Amzal, B., Bois, F., Parent, E., & Robert, C. (2006). Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association*, 101(474), 773-785.
- Atkinson, A., & Donev, A. (1992). *Optimum experimental designs*. Oxford University Press.
- Bardsley, W., Wood, R., & Melikhova, E. (1996). Optimal design: A computer program to study the best possible spacing of design points for model discrimination. *Computers & Chemistry*, 20, 145-157.
- Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3), 273-304.
- Cover, T., & Thomas, J. (1991). *Elements of information theory*. John Wiley & Sons, Inc.
- Ding, M., Rosner, G., & Müller, P. (2008). Bayesian optimal design for phase ii screening trials. *Biometrics*, 64, 886-894.
- Doucet, A., Freitas, N. de, & Gordon, N. (2001). *Sequential monte carlo methods in practice*. Springer.
- El-Gamal, M., & Palfrey, T. (1996). Economical experiments: Bayesian efficient experimental design. *International Journal of Game Theory*, 25, 495-517.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1996). *Markov chain monte carlo in practice*. Chapman & Hall.
- Haines, L., Perevozskaya, I., & Rosenberger, W. (2003). Bayesian optimal designs for phase i clinical trials. *Biometrics*, 59, 591-600.
- Heavens, A., Kitching, T., & Verde, L. (2007). On model selection forecasting, dark energy and modified gravity. *Monthly Notices of the Royal Astronomical Society*, 380(3), 1029-1035.
- Kirkpatrick, S., Gelatt, C., & Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220, 671-680.
- Kujala, J., & Lukka, T. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology*, 50(4), 369-389.
- Lesmes, L., Jeon, S.-T., Lu, Z.-L., & Doshier, B. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick tvc method. *Vision Research*, 46, 3160-3176.
- Müller, P., Sanso, B., & De Iorio, M. (2004). Optimal bayesian design by inhomogeneous markov chain simulation. *Journal of the American Statistical Association*, 99(467), 788-798.
- Myung, J. I., & Pitt, M. (in press). Optimal experimental design for model discrimination. *Psychological Review*.
- Nelson, J. (2005). Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, 112(4), 979-999.
- Opfer, J., & Siegler, R. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, 55, 165-195.
- Rubin, D., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology*, 25(5), 1161-1176.
- Rubin, D., & Wenzel, A. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103(4), 734-760.
- Tang, Y., Cavagnaro, D. R., Myung, J. I., & Pitt, M. A. (2009). *Adaptive design optimization for developmental psychology experiments*. (Presented at The 42nd Annual Meeting of the Society for Mathematical Psychology)